A simple GCV method of span selection for periodogram smoothing

By HERNANDO C. OMBAO[*], JONATHAN A. RAZ[†], ROBERT L. STRAWDERMAN[‡] and RAINER VON SACHS[§] [¶]

November 30, 1999

Abstract

A consistent estimator for the spectral density of a stationary random process can be obtained by smoothing the periodograms across frequency. An important component of smoothing is the choice of the span. Lee (1997) proposed a span selector that was claimed to be unbiased for the mean squared error. However, the symmetry property of the periodograms was overlooked and consequently this claim of unbiasedness is incorrect. In this paper, we propose a span selector originally developed for use in fitting generalized additive models. It is a form of the generalized cross validation function and is derived from the gamma deviance. Thus, it utilizes the approximate full likelihood of periodograms, which asymptotically behave like independently distributed gamma random variables. The proposed span selector is simple and easy to implement. Furthermore, unlike most plug-in methods, it does not need an estimate of any other unknown quantity or functional. We show finite sample simulation results that suggest that the proposed method is generally better than its competitor given in Lee (1997).

Some key words: Bandwidth selection; Deviance; Generalized additive model; Generalized cross-validation; Smoothed Periodograms; Spectrum estimation; Stationary time series.

[*]Departments of Statistics and Psychiatry, University of Pittsburgh, Pittsburgh, PA 15260-0001, U.S.A. ombao@stat.pitt.edu

[†]Division of Biostatistics, University of Pennsylvania, Philadelphia, PA 19104-6021, U.S.A. jraz@cceb.upenn.edu

[‡]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, U.S.A. strawder@sph.umich.edu

[§]Institut de Statistique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium rvs@stat.ucl.ac.be

[¶]Financial support from the contract "Projet d'Actions de Recherche Concertées" no. 98/03-217 of the Belgian Government is gratefully acknoledged.

1 Introduction

Let $X_t,\ t = 0, \pm1, \pm2\ldots$, be a zero mean stationary time series with covariance function $\gamma(\tau) = \mathrm{E}\{X_t X_{t+\tau}\},\quad \tau = 0, \pm1, \pm2, \ldots$. The second-order properties of $\{X_t\}$ are equivalently described by the spectrum

$$f(\lambda) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)\exp(-i2\pi\lambda\tau) \quad \text{where the frequency } \lambda \in [0,1]. \tag{1}$$

Since $f(\lambda)$ is symmetric about $\lambda = 1/2$, it suffices to estimate $f(\lambda)$ only on the interval $[0, 1/2]$.

One parametric method for estimating the spectrum is 'autoregressive spectral estimation' (Akaike, 1969). This method assumes that the stationary time series $\{X_0, \ldots, X_{T-1}\}$ is generated by some general linear process that can be adequately approximated by an autoregressive process of finite order p, denoted AR(p). A zero-mean stationary process $\{X_t\}$ is defined to be AR(p) if it can be written as $X_t = \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p} + \epsilon_t$ where ϵ_t's are independent, have mean zero and variance σ^2. (Details on the general linear process can be found in Priestley, 1981, p. 141) After fitting an AR(p) model to the time series, an estimate of the spectrum is given as

$$\hat{f}(\lambda) \quad = \quad \hat{\sigma^2}/\Big|1 - \hat{\beta}_1\exp(-i2\pi\lambda) - \ldots - \hat{\beta}_p\exp(-i2\pi p\lambda)\Big|.^2$$

The order p of the AR process can be selected from the data by using the Akaike information criterion (Priestley, 1981, §5.4.5, §7.8).

Various non-parametric methods for estimating the spectrum use the periodograms which are sample analogues of the spectrum. The periodogram at frequency $\lambda_k = k/T$ computed from a time series $\{X_0, \ldots, X_{T-1}\}$ is

$$\begin{aligned} I_k &= T^{-1}\Big|\sum_{t=0}^{T-1} X_t \exp\left(i2\pi kt/T\right)\Big|^2, \quad k = 0, \ldots, T-1. \\ &= T^{-1}|\sum_{t=0}^{T-1} X_t \cos(2\pi kt/T)|^2 \ + \ T^{-1}|\sum_{t=0}^{T-1} X_t \sin(2\pi kt/T)|^2. \end{aligned} \tag{2}$$

From Equation (2), I_k is essentially the squared correlation between the time series and the sine and cosine waves oscillating at frequency $\lambda_k = k/T$. The periodogram is known to be an asymptotically unbiased estimator of the spectrum. However, the I_k's are inconsistent estimators as they exhibit an erratic behavior even when the underlying spectrum of the process is smooth (Brillinger, 1981, pp. 120-125).

The traditional method for obtaining consistent estimators of the spectrum is to smooth the periodograms over frequency using a kernel smoother (Brillinger, 1981, p. 149). Let f_k be the value of the spectrum at frequency $\lambda_k = k/T$. A smoothed periodogram estimator of f_k, denoted $\widehat{f}_k$, is of the form

$$\widehat{f}_{k,p} = \sum_{j=-p}^{j=p} W_{p,j} I_{k+j}, \;\; k = 0, 1, \ldots, T/2 \tag{3}$$

where $2p + 1$ is the smoothing span and $W_{p,j}$ are non-negative weights that satisfy, for any fixed p, the following conditions:

$$W_{p,j} = W_{p,-j} \;\; \text{for} \;\; j = 1, \ldots, p, \;\; \text{and} \;\; \sum_{j=-p}^{j=p} W_{p,j} = 1. \tag{4}$$

For $\widehat{f}_{k,p}$ to be consistent, we must satisfy the conditions that $p = p_T \to \infty$ and $p_T/T \to 0$ as $T \to \infty$ (Brillinger, 1981, p.146). Generally, the weights are chosen so that $W_{p,0}$ is a decreasing function of p. There are various weights that satisfy the requirements in Equation (4). The form of the weights $W_{p,j}$, however, is only of secondary importance to the value of the span or bandwidth (Priestley, 1981, pp. 449 and 463).

The subject of span selection is well studied in non-parametric kernel density estimation. A bandwidth selection method is classified generally as either a “classical” or “plug-in type” method (Loader, 1999). The classical methods include those that are based on cross-validation, Akaike information criterion and other similar criteria. Plug-in type methods are those that use a pilot bandwidth that, in turn, is used to compute a pilot estimate of another unknown quantity. The application of plug-in type methods in the regression estimation setup is given in Gasser, Kneip and Köhler (1991). Plug-in type methods have been presented by its proponents to be superior to the classical methods; see Gasser, Kneip and Köhler (1991) and Hall, Sheather, Jones and Marron (1991). However, plug-in type methods are largely dependent on the arbitrary specification of pilot bandwidths. Therefore, they can give seriously biased estimates when the specification is wrong. Despite claims of its superiority in asymptotic studies, plug-in based estimates are also outperformed by their own pilot estimates (Loader, 1999).

Automatic span selection of kernel smooothers in spectral estimation has not received the same attention as in density estimation. Bühlmann (1996) proposed a plug-in type method for selecting spans that are allowed to change spatially according to the smoothness of the spectrum. It is expected to preserve significant peaks when smoothing periodograms and hence can perform better than a global span selector when the underlying spectrum is not smooth. Being a plug-in type method, however, the Bühlmann method can be sensitive to the choice of the pilot bandwidth. Recently, Lee (1997) proposed a span selector that is based on unbiased estimation of the mean squared error between the smoothed periodogram and the true spectrum. The Lee method does not require a pilot estimate of an unknown quantity. Unlike the Bühlmann method, the Lee method chooses a global rather than a local span. Thus, it is expected to work well when the spectrum is smooth globally. One problem with the Lee method, however, is that it assumes that all the periodograms $I_0, \ldots, I_{T-1}$ are independent. This is clearly not true (Brillinger, 1981, Theorem 5.2.6) and this assumption has led to an incorrect claim that the proposed span selector is unbiased for the mean squared error.

In this paper, we propose a simple alternative method in span selection for periodogram smoothing. The proposed method is based on a procedure originally developed for fitting generalized additive models (GAMs) (Hastie and Tibshirani, 1990). Our method employs the deviance function (a measure of deviation between the smoothed spectral estimate and the raw periodogram) that is appropriate for the asymptotic distribution of the periodogram. The proposed procedure for span selection has also been used as part of an automatic method for analysis of nonstationary time series (Ombao, Raz, von Sachs and Mallow, 1999). The rest of this paper is organized as follows. In Section 2, we give some background on the asymptotic behavior of periodograms and briefly describe the method in Lee (1997). In Section 3, we introduce the proposed alternative method. Finally, in Section 4, we provide simulation results comparing the proposed method and the Lee method. A bias-corrected version of the Lee span selector is presented in the Appendix.

2 Some background

Let $X_0, \ldots, X_{T-1}$ be a zero-mean stationary time series of length T where T is even. According to Theorem 5.2.6 in Brillinger (1981): when (i) all the moments of X_t exist; (ii.) $\sum |\gamma(\tau)| < \infty$; and (iii.) T is large, then

$$I_j/f_j \sim \epsilon_j, \; j = 0, 1, \ldots, T/2,$$

where ϵ_0 and $\epsilon_{T/2}$ are approximately distributed as χ_1^2 random variables and ϵ_1, $\ldots$, $\epsilon_{T/2-1}$ are approximately distributed as $\chi_2^2/2$ random variables. Furthermore, $\epsilon_0, \epsilon_1, \ldots, \epsilon_{T/2}$ are approximately independent.

As remarked in the Introduction, Lee (1997) recently proposed a method for selecting the span to be used in smoothing the periodograms. Following Pawitan and O'Sullivan (1994), Lee (1997) made the simplifying assumption that $\epsilon_0, \epsilon_1, \ldots, \epsilon_{T/2}$ are *each* exactly distributed as independent $\chi_2^2/2$ random variables. Note in particular that the distributions of ϵ_0 and $\epsilon_{T/2}$ are assumed to be the same as the other ϵ_j's, despite the fact that they are not. Pawitan and O'Sullivan (1994) claim that the difference between using the distributions χ_1^2 and $\chi_2^2/2$ at frequencies $\lambda_k = 0, 1/2$ is negligible asymptotically. In handling smoothing at the edges, i.e., when λ_k is close to 0 or 1, the Lee method used periodic smoothing. This is a sensible approach because both the spectrum and periodograms are periodic (the spectrum has period 1 and the periodograms T); for example, note that $I_{-1} = I_{T-1}$, $I_{-2} = I_{T-2}$, $\ldots$, and $I_{T+1} = I_1$, $I_{T+2} = I_2$ and so on.

An equivalent formulation of the Lee model is

$$I_j \quad \sim \quad \text{independent Gamma}(\alpha_j = 1, \beta_j = f_j), \quad j = 0, 1, \ldots, T-1, \tag{5}$$

where the probability density function of a Gamma(α, β) random variable is defined in Casella and Berger (1990, p. 627). Under (5), the Lee method chooses the span that minimizes

$$\widehat{R}(p) = \frac{RSS(p)}{T} - \frac{(1 - 2W_{p,0})}{2T} \sum_{j=0}^{T-1} I_j^2 \tag{6}$$

where

$$RSS(p) = \sum_{j=0}^{T-1} \left(I_j - \widehat{f}_{j,p}\right)^2 . \tag{7}$$

With only a slight modification to its last term, the criterion function (6) can be viewed as the generalized cross-validated (GCV) deviance function in a generalized additive model with normally distributed error terms; see Hastie and Tibshirani (1986, §6.9). Formula (6) is exactly the normal deviance function plus a correction intended to account for the bias of the usual cross-validated Gaussian deviance as an estimator of the $\mathcal{L}_2$-risk function

$$R_2(p) \quad = \quad T^{-1} \left\{ \sum_{j=0}^{T-1} E\left(f_j - \widehat{f}_{j,p}\right)^2 \right\} . \tag{8}$$

Lee claimed that (6) is an unbiased estimator of (8). This claim is based on the assertion that the periodograms $I_0, I_1, \ldots, I_{T-1}$ are independent. This assumption is incorrect because the periodograms are symmetric about $\lambda_k = 1/2$. Consequently, once $I_0, \ldots, I_{T/2}$ are specified, the remaning I_k's are completely determined. Moreover, the periodograms are periodic and are symmetric about $\lambda_k = 0$. Thus, $I_{-1} = I_1 = I_{T-1}$, $I_{-2} = I_2 = I_{T-2}$, etc .... Thus, one needs to be careful when smoothing at the edges especially when the smoothing window has a span that is "sufficiently large" so that it covers identical periodograms. The correct formula for the bias of the Lee span selector (6) as an estimator of (8) is derived in the Appendix. We give the basic result here. The bias for a fixed T, denoted $B_T(p)$, is

$$B_T(p) = \begin{cases} \left[-2\sum_{\ell=1}^{p/2} W_{p,2\ell}(f_\ell^2 + f_{T/2-\ell}^2) + (1 - 2W_{p,0})(f_0^2 + f_{T/2}^2)\right]/T, & p = 2, 4, \ldots \\ B_T(p-1), & p = 3, 5, \ldots \end{cases}$$

The order of the magnitude of the bias can be easily obtained when the weights chosen are such that $W_{p,j} = 1/(2p+1)$ for $j = -p, \ldots, p$ and $W_{p,j} = 0$, otherwise. For this choice of weights, the order of the magnitude of the bias is $O(1/T)$.

Though (6) represents squared error loss, the criterion has another potential drawback. In particular, as indicated above, the use of this criterion corresponds to selecting the smoothing parameter in a homoscedastic additive model with normal errors. Since the errors are heteroscedastic (the variance of I_j depends on the mean f_j), this casts some doubt on its appropriateness as a fitting criterion in this context, particularly with regard to efficiency. More generally, the problem of selecting the smoothing parameter can be viewed as fitting a generalized additive model in which the predictor variables are given by the frequency. This suggests that a more appropriate GCV criterion can be formulated based on a suitably constructed deviance function. We explore this idea further in Section 3.

3 The Gamma GCV Method

We propose a simple method of selecting the smoothing parameter using a procedure developed for GAMs. In line with Brillinger (1981, Theorem 5.2.6), consider the model

$$\begin{aligned} I_j &\sim \text{Gamma}(\alpha_j = 1, \beta_j = f_j), \quad j = 1, \ldots, T/2 - 1, \\ I_j &\sim \text{Gamma}(\alpha_j = 1/2, \beta_j = 2f_j), \quad j = 0, T/2, \end{aligned}$$

where $I_0, \ldots, I_{T/2}$ are independent. Under this model $E(I_j) = f_j$ for $j = 0, \ldots, T/2$, and the problem of spectral estimation can be considered in terms of a GAM with a gamma distribution, the identity link, and one predictor variable (frequency). Hastie and Tibshirani (1990, §6.9) discuss a number of different methods for automatic selection of the smoothing parameter in a GAM. The span selection methods discussed are the cross-validated deviance (analogous to cross-validation), generalized cross-validated deviance (analogous to generalized cross-validation) and the AIC statistic. All the span selection methods use a general form of the deviance rather than the usual sum of squared residuals.

In this paper, we select the smoothing parameter by minimizing the GCV deviance function

$$GCV(p) = \frac{M^{-1} \sum_{j=0}^{M-1} D(I_j,\ \widehat{f}_{p,j})}{\{\ 1 - \mathrm{tr}(H_p)/M\ \}^2}, \tag{9}$$

where $M = T/2 + 1$, $D(I_j, \widehat{f}_{j,p})$ is the deviance and H_p is the smoother matrix with smoothing parameter p. The deviance function appropriate for the present set-up (i.e., for independent gamma distributions with means governed by the true spectral values $f_j, j = 0, \ldots, T/2$) is

$$D(I_j, \widehat{f}_{j,p}) = 2\left\{-\log(I_j/\widehat{f}_{j,p}) + (I_j - \widehat{f}_{j,p})/\widehat{f}_{j,p}\right\}; \tag{10}$$

see McCullagh and Nelder (1989, pp. 33-34).

The denominator in (9), often referred to as the model degrees of freedom, can be expressed in terms of the weight at the center of the smoothing window: $\{1 - \mathrm{tr}(H_p)/M\}^2 = (1 - W_{p,0})^2$. Thus, the best span, $\widehat{p}_{GCV}$, is the minimizer of Gamma GCV function

$$GCV(p) = \frac{2M^{-1} \sum_{j=0}^{M-1} \left\{I_j/\widehat{f}_{j,p} - \log(I_j/\widehat{f}_{j,p}) - 1\right\}}{(1 - W_{p,0})^2}. \tag{11}$$

The Gamma GCV function above suggests that $W_{p,0}$ cannot be 1. Thus, the best p selected by the method must be ≥ 1.

The Gamma GCV method is conceptually simple and is easy to implement. The bias correction inherent in (11) is computationally straightforward and simpler than that employed in Lee (1997). The Gamma GCV method is also based on the full (asymptotic) likelihood because it uses the deviance that is appropriate for the asymptotic distribution of the periodograms. Hence, it is based on a deviance function that is arguably more relevant than that considered in Lee (1997). It also has an appealing interpretation since, as described in Hastie and Tibshirani (1986, §6.8.3), (11), evaluated at its minimizer $\widehat{p}_{GCV}$, may be viewed as an approximately unbiased estimate of the prediction error $E[\sum_j D(I_j, \hat{f}_{j,\hat{p}})/M]$. Finally, unlike most plug-in methods, it does not require a pilot estimate of any unknown quantities, and hence does not suffer from the problems that can result from misspecifying such pilot estimators.

4 Numerical Experiments

4.1 The Set Up

A simulation study was conducted to compare the performance of the Lee method and Gamma GCV method using six different ARMA (α, β) processes. We define a random process $\{X_t\}$ to be ARMA(α, β) if it is generated by the equation $X_t + a_1 X_{t-1} + \ldots a_\alpha X_{t-\alpha} = \xi_t + b_1 \xi_{t-1} + \ldots + b_\beta \xi_{t-\beta}$ where ξ_t are independent $N(0,1)$ random variables. In example 1, we used ARMA(3,0) with $a_1 = -1.5, a_2 = 0.7, a_3 = -0.1$. In example 2, ARMA(0,4) with $b_1 = -0.3, b_2 = -0.6, b_3 = -0.3, b_4 = 0.6$. In example 3, ARMA(3,0) with $a_1 = 0.9, a_2 = 0.8, a_3 = 0.6$. In example 4, ARMA(0,3) with $b_1 = 0.9, b_2 = 0.8, b_3 = 0.6$. In example 5, ARMA(1,0) with $a_1 = 0.05$. In example 6, ARMA(0,0) or the white noise process. Examples 1 to 4 were presented in Lee (1997). We also used the Daniell smoother whose weights are defined by

$$W_{p,j} = \begin{cases} 1/(2p), & \text{if } \; -(p-1) \le j \le (p-1) \\ 1/(4p), & \text{if } \; j \in \{-p, p\} \end{cases}$$

The simulation study was done using the S-PLUS Statistical Software. In each ARMA example, 1000 independent time series data sets, each of length $T = 1024$, were generated.

4.2 The Criteria

For the n-th data set generated in each ARMA example ($n = 1, \ldots, 1000$), we obtained the spans selected by the Lee and the Gamma GCV methods, denoted by $\widehat{p}_{L,n}$ and $\widehat{p}_{GCV,n}$ respectively. Next, we computed the smoothed spectral estimates using the Lee and the Gamma GCV spans. We then computed the Lee and Gamma GCV average loss (over 1000 generated data sets) under the four criteria that we will describe below. These four criteria give some measure of closeness of the spectral estimates to the true spectrum.

The first criterion is the **squared error**, or $\mathcal{L}_2$, loss, a popular measure for evaluating estimators. The $\mathcal{L}_2$ loss for an estimate $\widehat{f}(\widehat{p}) = \{\widehat{f}_{k,\widehat{p}},\ k = 0, \ldots, T/2\}$ is defined by

$$\widetilde{R_2}(\widehat{p}) \;=\; T^{-1} \sum_{k=0}^{T-1} (f_k - \widehat{f}_{k,\widehat{p}})^2. \tag{12}$$

We computed the squared error loss for the Lee and Gamma GCV spectral estimates for the n-th data set. These are denoted by $\widetilde{R_2}(\widehat{p}_{L,n})$ and $\widetilde{R_2}(\widehat{p}_{GCV,n})$ respectively. The average squared error loss for the Lee and Gamma GCV methods (out of 1000 data sets) are given in Table 1.

As a second criterion, we considered the **gamma deviance** as the loss function for evaluating $\widehat{f}(\widehat{p})$. The gamma deviance for $\widehat{f}(\widehat{p})$ is given by

$$\widetilde{G}(\widehat{p}) \;=\; 2M^{-1} \sum_{k=0}^{M-1} \left\{ \widehat{f}_{k,\widehat{p}}/f_k - \log(\widehat{f}_{k,\widehat{p}}\,/f_k) - 1 \right\}. \tag{13}$$

We computed the gamma deviance for the Lee and Gamma GCV estimates, respectively denoted by $\widetilde{G}(\widehat{p}_{L,n})$ and $\widetilde{G}(\widehat{p}_{GCV,n})$. The average gamma deviance for the Lee and the Gamma GCV spectral estimates are given in Table 2. The simulated average gamma deviance approximates the asymptotic Kullback-Leibler distance between the estimated spectrum and the true spectrum.

The above two "designer criteria" are considered for obvious reasons. It is reasonable to expect that the Lee method (which selects the span by minimizing an estimate of the mean squared error) should outperform the Gamma GCV method under the $\mathcal{L}_2$ loss criterion. Similarly, it is fair to expect the Gamma GCV method to beat the Lee method under the gamma deviance loss criterion. To explore the performance of these estimators to alternative criteria, we also considered two other norms. The first alternative criterion is the **absolute error**, or $\mathcal{L}_1$, loss defined by

$$\widetilde{R_1}(\widehat{p}) \;=\; T^{-1} \sum_{k=0}^{T-1} \mid f_k - \widehat{f}_{k,\widehat{p}} \mid. \tag{14}$$

The second alternative criterion is the **supremum error**, or $\mathcal{L}_\infty$, loss defined by

$$\widetilde{R_\infty}(\widehat{p}) \;=\; \sup_k \mid f_k - \widehat{f}_{k,\widehat{p}} \mid. \tag{15}$$

The $\mathcal{L}_1$ loss for the Lee and Gamma GCV spectral estimates are respectively denoted $\widetilde{R_1}(\widehat{p}_{L,n})$ and $\widetilde{R_1}(\widehat{p}_{GCV,n})$; similarly, for the $\mathcal{L}_\infty$ loss, we use $\widetilde{R_\infty}(\widehat{p}_{L,n})$ and $\widetilde{R_\infty}(\widehat{p}_{GCV,n})$. The average $\mathcal{L}_1$ and $\mathcal{L}_\infty$ losses for the Lee and Gamma GCV estimates over 1000 generated data sets are respectively summarized in Tables 3 and 4.

4.3 Discussion of Simulation Results

Under the squared error criterion, the average loss and the standard errors for the Gamma GCV and the Lee methods are quite similar. This result was somewhat unexpected because the Lee method is developed for $\mathcal{L}_2$. Thus, it was expected to perform better than the Gamma GCV method. The apparent failure of the Lee method to dominate under the $\mathcal{L}_2$ loss criterion might be due to the fact that the Lee span selector given in Equation (6) is not truly unbiased for the $\mathcal{L}_2$ risk. Under the gamma deviance criterion, the Gamma GCV method substantially outperformed the Lee method in every example. This result was expected given the risk criterion used to select the smoothing parameter. Under $\mathcal{L}_1$, the Gamma GCV method had smaller average loss than the Lee method in all examples; the same is true for $\mathcal{L}_\infty$ except for Examples 2 and 4. Moreover, the study shows that the standard errors of the mean loss of the Gamma GCV method, under each criterion, are generally smaller than that obtained by the Lee method. This suggests that the Gamma GCV method may give more stable and less variable spectral estimates. This may be at least partly attributable to the fact that least squares, which is the basic criterion upon which Lee's procedure is based, leads to consistent but inefficient estimators under heteroscedasticity.

ACKNOWLEDGMENTS

Hernando Ombao and Jonathan Raz were supported by NIMH Grant R29 MH51310. Rainer von Sachs gratefully acknowledges financial support from the contract "Projet d'Actions de Recherche Concertes" no. 98/03-217 of the Belgian Government.

5 Appendix

5.1 Bias of the Lee span selector

In the appendix, we will evaluate the order of the magnitude of the bias of the Lee span selector, $\hat{R}(p)$ (6), as an estimator of the mean square error $R_2(p)$ (8). We then propose a bias-corrected version. We start by evaluating $E(RSS(p))$.

$$\begin{aligned} E(RSS(p)) &= \sum_{j=0}^{T-1} E(I_j - \hat{f}_{j,p})^2 \\ &= E(I_0 - \hat{f}_{0,p})^2 + E(I_{T/2} - \hat{f}_{T/2,p})^2 + 2\sum_{j=1}^{T/2-1} E(I_j - \hat{f}_{j,p})^2. \quad (16) \end{aligned}$$

Equation (16) follows because the periodograms are symmetric about the frequency index $j = T/2$. We follow the notation that ϵ_ℓ, $\ell = 0, \ldots, T/2$, are independently distributed random variables; $\epsilon_1, \ldots, \epsilon_{T/2-1}$ are identically distributed as $\chi^2_2/2$ random variables; ϵ_0 and $\epsilon_{T/2}$ are distributed as χ^2_1 random variables.

We first simplify the first term (frequency index $j = 0$) in (16).

$$E(I_0 - \hat{f}_{0,p})^2 \quad = \quad E(I_0^2 - 2I_0\hat{f}_{0,p} + \hat{f}_{0,p}^2) \tag{17}$$

We compute the cross term in the above equation.

$$\begin{aligned} E(I_0\hat{f}_{0,p}) &= E(W_{p,0}I_0^2 + \sum_{k\neq 0} W_{p,k}I_0I_{0+k}) \\ &= E(W_{p,0}f_0^2\epsilon_0^2 + \sum_{k\neq 0} W_{p,k}f_0f_{0+k}\epsilon_0\epsilon_{0+k}) \\ &= 3W_{p,0}f_0^2 + \sum_{k\neq 0} W_{p,k}f_0f_{0+k}. \end{aligned} \tag{18}$$

Equation (18) follows because $E(\epsilon_0^2) = 3$ and $E(\epsilon_0\epsilon_{0+k}) = 1$ unless $k = 0, \pm T$. We already know that $k \neq 0$. Moreover, k cannot be equal to $\pm T$ because $-p \leq k \leq p$ and $|p| < T$. Thus,

$$E(I_0\hat{f}_{0,p}) \quad = \quad 2W_{p,0}f_0^2 + f_0E(\hat{f}_{0,p}).$$

Equation (17) can be written as

$$E(I_0 - \hat{f}_{0,p})^2 = E(f_0 - \hat{f}_{0,p})^2 + 2(1 - 2W_{p,0})f_0^2. \tag{19}$$

Similarly, we can derive expected squared error for frequency index $j = T/2$ to be

$$E(I_{T/2} - \hat{f}_{T/2,p})^2 = E(f_{T/2} - \hat{f}_{T/2,p})^2 + 2(1 - 2W_{p,0})f_{T/2}^2. \tag{20}$$

Next, we shall compute the last term of (16). For any frequency index $j = 1, \ldots, T/2 - 1$,

$$E(I_j - \hat{f}_{j,p})^2 \quad = \quad E(I_j^2 - 2I_j\hat{f}_{j,p} + \hat{f}_{j,p}^2) \tag{21}$$

We already know that $E(I_j^2) = 2f_j^2$. We now work with the cross term in the above equation.

$$\begin{aligned} E(I_j\hat{f}_{j,p}) &= E(W_{p,0}I_j^2 + I_j\sum_{k\neq 0} W_{p,k}I_{j+k}) \\ &= 2W_{p,0}f_j^2 + \sum_{k\neq 0} E(W_{p,k}f_jf_{j+k}\epsilon_j\epsilon_{j+k}). \end{aligned} \tag{22}$$

We note that $E(\epsilon_j \epsilon_{j+k}) = 1$ whenever $2j + k \neq 0, \pm T$. Otherwise, $E(\epsilon_j \epsilon_{j+k}) = 2$. Thus, in Equation (22), we arrive at $E(I_j \hat{f}_{j,p}) = 2W_{p,0} f_j^2 + \sum_{k \neq 0} W_{p,k} f_j f_{j+k} + \Delta(p, j)$, where $\Delta(p, j) = \sum_{k \in \mathcal{J}} W_{p,k} f_j f_{j+k}$; and $\mathcal{J} = \{-p, \ldots, p;\ k \neq 0;\ 2j + k = 0, \pm T\}$. Thus,

$$E(I_j \hat{f}_{j,p}) \quad = \quad W_{p,0} f_j^2 + f_j E(\hat{f}_{j,p}) + \Delta(p, j)$$

Equation (21) can now be written as

$$E(I_j - \hat{f}_{j,p})^2 \quad = \quad E(f_j - \hat{f}_{j,p})^2 + (1 - 2W_{p,0}) f_j^2 - 2\Delta(p, j) \tag{23}$$

By substituting results in Equations (19), (20) and (23) into Equation (16), we arrive at

$$E\{RSS(p)\} = E \sum_{j=0}^{T-1} (f_j - \hat{f}_{j,p})^2 + (1 - 2W_{p,0}) \sum_{j=0}^{T-1} f_j^2 - 2 \sum_{j=0}^{T-1} \Delta(p, j) + \Delta_0(p) \tag{24}$$

where $\Delta_0(p) = (1 - 2W_{p,0})(f_0^2 + f_{T/2}^2)$. Using Equations (6) and (8), we have

$$E\{\hat{R}(p)\} \quad = \quad R_2(p) + B_T(p) \tag{25}$$

where $B_T(p) = [-2 \sum_j \Delta(p, j) + \Delta_0(p)]/T$.

It is instructive to consider the bias $B_T(p)$ at specific values of p. First, we set $p = 2$. The values of $\Delta(2, j)$ are computed to be: 0 when $j = 0$; $W_{2,-2} f_1^2$ when $j = 1$; 0 when $j = 2, \ldots, T/2 - 2$; $W_{2,2} f_{T/2-1}^2$ when $j = T/2 - 1$; $W_{2,-2} f_{T/2+1}^2$ when $j = T/2 + 1$; 0 when $j = T/2 + 2, \ldots, T - 2$; and $W_{2,2} f_{T-1}^2$ when $j = T - 1$. The computation of the bias can be simplified when we use the symmetry property of the spectrum $f_1 = f_{T-1}$ and $f_{T/2-1} = f_{T/2+1}$; and the weights $W_{p,j} = W_{p,-j}$. Thus, at $p = 2$, the bias is equal to

$$B_T(2) = \left[-2W_{2,2}(f_1^2 + f_{T/2-1}^2) + (1 - 2W_{p,0})(f_0^2 + f_{T/2}^2)\right]/T.$$

When $p = 3$, it is easy to arrive at $B_T(3) = B_T(2)$. We now give a general form of the bias $B_T(p)$:

$$B_T(p) = \begin{cases} \left[-2 \sum_{\ell=1}^{p/2} W_{p,2\ell}(f_\ell^2 + f_{T/2-\ell}^2) + (1 - 2W_{p,0})(f_0^2 + f_{T/2}^2)\right]/T, & p = 2, 4, \ldots \\ B_T(p - 1), & p = 3, 5, \ldots \end{cases}$$

We now evaluate the order of magnitude of the bias for the case where we use the boxcar window $W_{p,j} = 1/(2p + 1)$ for $j = -p, \ldots, p$ and $W_{p,j} = 0$, otherwise. Furthermore, if we assume that there exists $K > 0$ so that $f_j < K$ for all j then $|B(p)| \leq (1/T)\{1/(2p+1)\}pK^2 + (1/T)\{1 - 2/(2p+1)\}(2K^2)$. Thus, the magnitude of the bias is of the order $O(1/T)$.

5.2 Bias-corrected version of the Lee span selector

Note that $E(I_0^2/3) = f_0^2$; $E(I_{T/2}^2/3) = f_{T/2}^2$; and $E(I_j^2/2) = f_j^2$ for all $j = 1, \ldots, T/2 - 1$. Hence, an unbiased estimator of $B_T(p)$ is

$$\hat{B}_T(p) = \begin{cases} \left[-\sum_{\ell=1}^{p/2} W_{p,2\ell}(I_\ell^2 + I_{T/2-\ell}^2) + (1 - 2W_{p,0})(I_0^2 + I_{T/2}^2)/3 \right] /T, & p = 2, 4, \ldots \\ \hat{B}_T(p-1), & p = 3, 5, \ldots \end{cases} \tag{26}$$

A bias-corrected version of the Lee span selector for a fixed T, denoted $\hat{R}_{BC}(p)$, is

$$\hat{R}_{BC}(p) = \hat{R}(p) + \hat{B}(p). \tag{27}$$

References

Akaike, H. (1969). Power spectrum estimation through autoregressive fitting. *Annals of the Institute of Statistical Mathematics*, **21**, 407-419.

Brillinger, D. (1981). *Time Series: Data Analysis and Theory,* Oakland, California: Holden-Day.

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal of Time Series Analysis,* **17**, 247-270.

Casella, G. and Berger, R. (1990). *Statistical Inference,* Belmont, California: Duxbury Press.

Gasser, T., Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, **86**, 643-652.

Hall, P., Sheather, S., Jones, M. and Marron, S. (1992). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, **91**, 401-407.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* London: Chapman and Hall.

Lee, T. (1997). A simple span selector for periodogram smoothing. *Biometrika,* **84**, 965-9.

Loader, C. (1999). Bandwidth selection: classical or plug in?. *Annals of Statistics*, **27**, 415-438.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models, Second Edition*. London: Chapman and Hall.

Ombao, H., Raz, J., von Sachs, R. and Mallow, M. (1999). Automatic statistical analysis of bivariate non-stationary time Series. Discussion Paper 99/08, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Pawitan, Y. and O'Sullivan, F. (1994). Nonparametric spectral density estimation using penalized Whittle likelihood. *Journal of the American Statistical Association,* **89**, 600-610.

Priestley, M. (1981). *Spectral Analysis and Time Series.* London: Academic Press Limited.

Table 1: Average of squared error loss for the Lee and Gamma GCV methods: $\tilde{R}_2(\widehat{p}_L)$ and $\tilde{R}_2(\widehat{p}_{GCV})$, computed from 1000 simulated time series data sets. Standard errors of the mean are inside the parentheses.

Example	Lee	Gamma GCV
1	1459.717	1465.717
	(57.210)	(57.430)
2	20.944	20.941
	(0.806)	(0.805)
3	39.449	39.449
	(1.575)	(1.575)
4	59.991	59.931
	(2.339)	(2.338)
5	4.595	4.596
	(0.157)	(0.157)
6	4.561	4.561
	(0.160)	(0.160)

Table 2: Average of the gamma deviance for the Lee and Gamma GCV methods: $\tilde{G}(\widehat{p}_L)$ and $\tilde{G}(\widehat{p}_{GCV})$, computed from 1000 simulated time series data sets. The standard errors of the means are inside the parentheses. The averages and standard errors are expressed in units of 10^{-3}.

Example	Lee	Gamma GCV
1	10.920	6.320
	(0.332)	(0.096)
2	17.750	10.530
	(0.295)	(0.119)
3	11.611	8.860
	(0.234)	(0.114)
4	19.360	8.737
	(0.397)	(0.116)
5	3.116	2.564
	(0.011)	(0.001)
6	2.917	2.392
	(0.100)	(0.008)

Table 3: Average of the absolute error loss for the Lee and Gamma GCV methods: $\tilde{R}_1(\widehat{p}_L)$ and $\tilde{R}_1(\widehat{p}_{GCV})$, computed from 1000 simulated time series data sets. The standard errors of the means are inside the parentheses. Values are expressed in units of 10^{-1}

Example	Lee	Gamma GCV
1	18.201	15.963
	(0.246)	(0.191)
2	27.102	26.720
	(0.002)	(0.002)
3	4.054	3.812
	(0.034)	(0.027)
4	3.886	3.651
	(0.048)	(0.035)
5	0.798	0.743
	(0.012)	(0.010)
6	0.769	0.712
	(0.012)	(0.011)

Table 4: Average of the $\mathcal{L}_\infty$ loss for the Lee and Gamma GCV methods: $\tilde{R_\infty}(\widehat{p}_L)$ and $\tilde{R_\infty}(\widehat{p}_{GCV})$, computed from 1000 simulated time series data sets. The standard errors of the means are inside the parentheses.

Example	Lee	Gamma GCV
1	33.037	25.039
	(0.736)	(0.339)
2	1.287	1.483
	(0.022)	(0.014)
3	3.789	3.342
	(0.057)	(0.034)
4	2.802	2.931
	(0.076)	(0.036)
5	0.195	0.170
	(0.004)	(0.003)
6	0.177	0.154
	(0.004)	(0.003)